面向事件的视频语义表示方法*

李旭晖　吴青峰
武汉大学信息管理学院　武汉 430072

摘　要：[目的/意义]视频内容正在影响着我国大量人口的信息生活，视频语义的良好表示是推动当前视频内容研究和视频应用服务向前发展的关键基础。现有的视频语义表示方法存在事件语义表示角度和粒度划分方式单一、缺少灵活的对象语义变化机制的问题，因此探究更有效的视频语义表示方法具有重要意义。[方法/过程]提出面向事件的视频语义表示方法。此方法考虑人的双向认知过程，可以根据不同用户背景和需求从不同角度解读和生成事件语义，并定义相应的语义对象和角色的变化机制。[结果/结论]面向事件的视频语义表示方法具有完整的语义表示框架，支持多角度的事件语义表示，可以灵活地进行属性级、对象级和事件级的语义拓展，能够表示更丰富的视频语义。

关键词：视频语义表示　多角度　语义拓展
分类号：G250
DOI：10.13266/j.issn.0252-3116.2020.10.011

1　引言

短视频、在线网课、手机直播、视频博客等基于视频媒介的内容正在影响我国大量人口的信息生活[1]。相较于电视时代和 PC 时代，移动互联网时代的视频内容的消费需求和相关研究正变得越来越精细化。视频内容研究最初只涉及视频的外部标注信息，之后开始关注视频中的颜色、运动轨迹等底层特征，而当下的重点则是基于视频语义特征的研究和应用。视频数据挖掘相关研究需要良好的视频语义模型作为建模基础，图书馆视频资源的高效组织和价值发挥有赖于合适的视频语义表示框架[2-3]，新型检索和推荐系统需要考虑视频语义才能从本质上提升视频内容的分发效率。

良好的视频语义表示方法是上述研究和应用中的基础关键。随着视频分析技术的发展和用户需求的精细化，视频语义表示研究不仅需要有效地包含语义对象和表示事件语义，也需要关注视频事件语义结构的设计、事件语义的可扩展性及相应的对象语义变化机制。

现有研究中的视频语义表示方法虽然具有一定的语义表示能力，但都存在事件语义表示角度和粒度划分方式单一、缺少灵活的对象语义变化机制等问题。

基于当前研究的现状，本文提出面向事件的视频语义表示方法，该方法遵循自底向上的语义描述过程，充分考虑用户的双向认知过程[4]，旨在提供支持事件语义的多角度表示及相应的多种粒度划分方式的视频语义表示方法。以该研究目的为核心，本文首先将会总结视频语义表示方法的研究现状和不足之处；然后，剖析当前的视频语义表示工作中亟待解决的几个关键问题；进而具体定义面向事件的视频语义表示方法，并论述本文方法对关键问题的解决，再以篮球比赛视频片段举例说明方法的有效性，并与现有相关方法比较，说明本文方法的创新性；最后总结全文，提出围绕该工作的未来研究方向。

2　相关研究

早期的视频语义表示采用的是基于标注的方法，其主要思想是将自然文本或结构数据组成的标注信息叠加在视频流中对应的视频序列上。这种方法可表示

* 本文系国家自然科学基金重大研究计划“大数据驱动的管理与决策研究”重点支持项目“基于知识关联的金融大数据价值分析、发现及协同创造机制”（项目编号：91646206）研究成果之一。
作者简介：李旭晖（ORCID：0000-0002-1155-3597），副教授，硕士生导师，E-mail：lixuhui@whu.edu.cn；吴青峰（ORCID：0000-0002-3967-3187），硕士研究生。
收稿日期：2019-12-04　修回日期：2020-02-23　本文起止页码：99-108　本文责任编辑：徐健

的语义对象有限，且标注之间无法互相关联，难以刻画视频中复杂的语义关系。它主要用于满足简单的基于关键字和属性的视频查询需求。随着视频资源的丰富和相关研究的发展，研究人员和用户对视频语义表示的需求也变得更加复杂。在视频数据挖掘领域，视频概念检测[5-6]、视频分类[7]、内容结构分析[8]、主题挖掘[9]、事件挖掘[10-11]等方面的研究都需要更有效的语义表示方法作为其前期建模基础，并为研究结果提供可解释性。在日益增长的新需求下，一方面，当前的视频语义表示相关研究已大多基于视频语义数据模型，是一种分层模型，底层和顶层分别对应原始视频数据流和视频语义信息。顶层的语义信息是通过对原始视频数据的语义抽象和映射得到的，基于映射机制的不同，模型可能具有不同种类和数量的中间层，例如语义对象层、事件场景层等。另一方面，自从领域事件[12]的概念提出以来，基于事件的语义表示也成为研究者们对视频语义表示研究的共识。事件是由一个或多个语义对象的特征、关系和背景信息等形成的较为完整和综合的语义信息单元。此后的视频语义表示相关研究虽然有不同的侧重，但基本上都体现了以事件为核心语义的分层模型的思想，本文的方法亦是以此为基础。

虽然有了共同的思想基础，但受制于视频分析技术水平或研究者所在领域的限制，许多相关研究对高层语义的表示较为局限。在王昊然等提出的基于图模型的足球视频语义表示方法[13]中，事件单元是由镜头和音频特征组成的。张静等则以行人运动特征来定义相关的事件模板，将运动轨迹映射为事件语义[14]。刘晓璐提出了安防视频的知识元模型[15]，将安防视频内容映射到视频基础信息、载体对象、安防事件 3 个方面。谢潇等定义了地理视频语义的多层次结构[16]，将地理视频语义抽象为相互关联的特征域、行为过程域、事件域 3 个层次。以上研究多集中于安防、地理、交通等领域，充分考虑了视频内容的底层特征和时空信息，并结合了专业领域知识，但缺乏对高层语义的支持，对事件语义的表示较为初级，也不具备通用性。因此本文的方法将重点关注视频中较为复杂的事件语义的表示，并尽量实现方法的通用性。

研究者们尝试提出了多种视频语义表示方法，旨在涵盖较为完整的视频底层特征信息，同时准确表达视频的高层语义信息。由 S. Adali 等提出的 AVIS[17]是较早引入高层语义信息的视频语义模型，AVIS 中明确定义了视频对象、事件、角色等语义概念，并将视频序列节点按时序包含关系形成树结构。但该模型忽略了语义对象的大部分底层特征，用户无法对对象语义进行自顶向下地扩展和补全。而 VIDEX 方法[18]中虽然集成了视频底层特征，但其对高层语义的结构设计比较简单，无法表示语义对象和事件间的复杂关联。鲍泓等提出了分层语义联想模型[19]，模型中使用了概念层次树表示抽象概念间的继承关系，能够有效地表示较为复杂的抽象概念，但未考虑事件语义结构的层次性。由 Y. Wang 提出的 THVDM[20]在概念层区分了对象和事件，预定义了一些事件语义结构，能够表示不同粒度的事件间的关联。但现有的表示方法都存在事件语义表示角度和粒度划分方式单一、缺少灵活的对象语义变化机制的问题。问题具体表现为：①事件语义表示角度和粒度划分方式单一。现有表示方法对事件的解读角度都是唯一的，而不同用户群体对视频语义的理解并不是统一的。以篮球比赛视频为例，教练可能从全局战术角度来解读，普通球迷可能从单一球员的表现来解读。事件语义的粒度划分也与解读角度有关，教练对整场比赛视频的事件粒度划分可能是依据整体战术的博弈过程来划分，普通球迷角度的事件粒度可能是依据单一球员的得分、犯规等具体行为来划分。②缺少灵活的对象语义变化机制。现有方法中的语义对象和参与事件的角色通常是静态的，无法很好地支持同一个语义对象在不同解读角度下的意义变化。对于可能参与不同粒度事件的语义对象，其对应的实例化角色的数量和意义的变化也需要具体的中间机制，现有方法对这类语义的变化未能提供良好的支持。因此本文提出面向事件的视频语义表示方法，关注事件语义的多角度表示及相应的多种粒度划分方式，并提供灵活的对象语义变化机制。

3 视频语义表示的关键问题

3.1 自底向上的描述过程

人对视频语义的认识是一个从底层物理特征到高层语义信息的自底向上的描述过程。一个符合实际认知过程的视频语义表示方法要能够有效且灵活地支持自底向上的描述过程。语义对象的表示是自底向上的描述过程的基础，在表示底层的语义对象时，需着重刻画语义对象与具体事件无关的特征，以保证同一个语义对象能够在不同事件中被重复调用。

在自底向上的描述过程中，表示高层的事件语义时，同一个语义对象在参与不同事件时会具有不同的语义信息。如篮球比赛视频中的“某运动员”对象就

可能分别以“进攻者”和“防守者”的语义参与“持球进攻”事件、“篮下防守”事件，因此可以引入专门表示语义对象的事件相关特征的中间角色。

3.2 视频事件的多角度解读

事件语义的表示是视频语义表示的核心，现有的视频语义表示方法缺少对事件的多角度解读的关注，它们忽略了人的认知行为的双向性特点。具体而言，在自下而上的描述过程中，人们可能会选择关注不同的语义对象或是同一语义对象的不同方面，因此可能会解读出不同的事件语义及相应的事件结构。例如在观看同一段篮球赛视频时，有的用户关注比分信息，有的用户关注某个球星的行为，有的用户关注裁判的行为，有的用户关注球队的战术配合，这些不同的关注点都可能是解读视频语义和划分事件结构的依据(见图 1)。因此，在视频语义表示中支持事件的多角度解读是非常重要的，语义表示方法中要能够表示各个角度的语义信息，并支持不同角度下的事件结构划分方式。

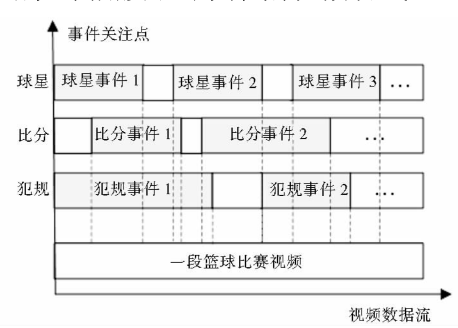


图 1 基于不同的关注点划分事件结构

3.3 视频事件的粒度划分

粒度是指视频事件被表示时的语义片段的大小。在上一个关键问题中提到了事件解读角度的选取会影响事件结构的划分，事件结构的划分主要体现为事件的粒度划分。事件的粒度划分既涉及事件之间的组合，也涉及相关事件角色的语义变化。事件之间的组合是指多个粒度较小、层次较低、语义单一的连续事件，可以作为子事件组合成粒度较大、层次较高、语义丰富的复合事件。比如一段监控视频中记录的连续的“追逐”“制服”“押送”事件可以组合成更大粒度的“抓捕”事件。在事件组合过程中，需要注意的是，子事件中涉及的语义对象在复合事件中会发生数量和语义上的变化，语义表示方法要能够支持这种变化，比如子事件“追逐”中的“被追逐者”角色在复合事件“抓捕”中就可演变为“犯人”角色。

3.4 自顶向下的语义补全

语义补全是指在已经形成的语义表示框架中填充更多语义信息。在自底向上的描述过程中，主要的认知工作是判断和确定语义对象及事件的存在和类型。而一旦确定了语义描述框架，通常就需要根据不同的需求，为框架中的语义对象或事件填充更多的语义信息，这些信息可以来自底层特征、背景信息等。例如，在一段篮球比赛视频中，先自底向上地根据运动员的时空关联等语义信息确定了“得分”事件，并赋予该事件中的语义对象“某运动员”在该事件中的角色为“得分手”。此时，除了“某运动员”固有的事件无关的属性外，可能还需要补全“得分手”角色的更多语义信息，比如“得分方式”是“投篮”“上篮”还是“罚球”。

3.5 对可扩展性和检索需求的支持

如前所述，不同用户可能对同一视频语义有不同的理解，因此良好的视频语义表示方法要具有灵活的可扩展性，要能够基于用户的不同关注点可扩展地生成语义，事件中涉及的对象和角色的语义也要能够随事件的语义变化进行扩展。

对检索需求的支持是视频语义表示方法在应用环节中发挥作用的重要能力，在当前丰富的视频内容和复杂的用户需求背景下，现有的基于关键字或底层特征的检索方法无法完全满足用户需求，而基于视频语义的检索方法是当前视频检索领域的研究和发展的关键，因此在视频语义表方法中考虑对多样化检索的支持是必要的。

4 视频语义表示方法

为解决上述视频语义表示中的关键问题，本文提出了面向事件的视频语义表示方法。本节将首先定义语义表示方法框架，并论述该方法解决上述关键问题的能力，再使用篮球比赛视频片段进行应用实例描述，最后对本文的方法与现有相关方法进行比较，说明本文方法的创新性。

4.1 面向事件的视频语义表示方法

本文提出了面向事件的视频语义表示方法，该方法的语义表示的逻辑框架如图 2 所示，语义表示框架中刻画了语义对象(Object)、角色(Role)和事件(Event)三类主体以及它们之间的关联方式。语义对象是该方法中语义表示的基础，角色由语义对象实例化得到，并参与具体事件的语义构建。对三类主体的具体定义如下：

(1)语义对象(Object)：语义对象是通过自动分析视频底层特征得到的具有初级语义的对象。所有语义对象都是独立于具体事件而存在的，使用元组将语义

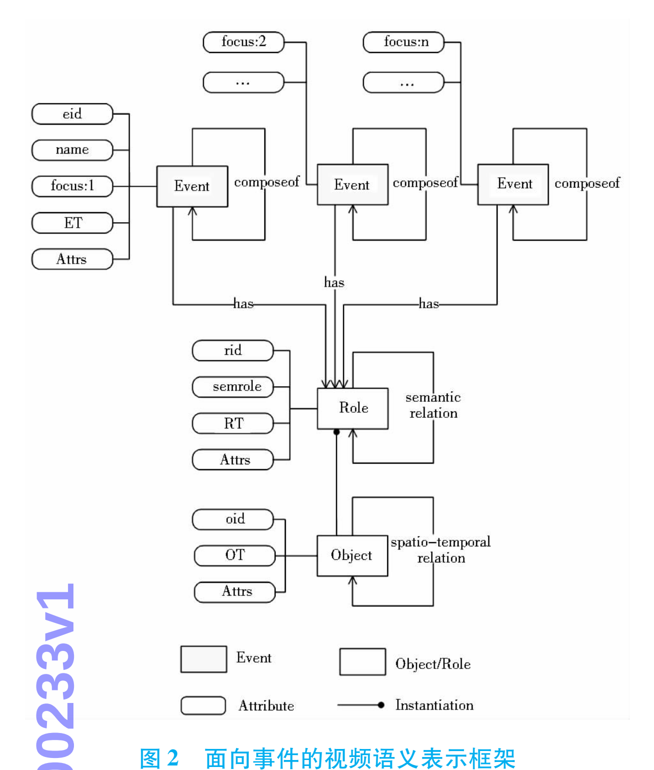


图 2 面向事件的视频语义表示框架

对象表示为 Object = {oid, OT, Attrs},其中:

① oid 是语义对象的唯一标识符。

② OT = {t_i,t_j}是语义对象在视频中出现的时间区间的起止时间记录。其中,ti 表示起始时间,tj 表示终止时间。

③ Attrs = {k_1: v_1,…, k_n: v_n}是可扩展的属性键值对。其中包含该对象的底层特征,如颜色、运动轨迹等;也包含其他事件无关的信息,如对象的命名等。可根据需要扩展或补全该对象的其他与事件无关的信息。

(2)角色(Role):角色是由语义对象在具体事件中实例化得到的。角色的语义信息是与具体事件相关的,使用元组将角色表示为 Role = {rid, semrole, RT, Attrs},其中:

① rid 是角色的唯一标识符。

② semrole 是角色的语义标签,表示角色在事件中扮演的语义角色类型。比如"抓捕"事件中的"警察""罪犯"。

③ RT = {t_i, t_j}是该角色在视频中的时间区间的起止时间记录。

④ Attrs = {k_1: v_1, …, k_n: v_n}是可扩展的属性键值对。其中的属性是结合相关语义对象的特征与其参与的事件语义得到的角色信息,可根据需要进行扩展或补全与事件相关的其他信息。

(3)事件(Event):事件是综合一个或多个有意义的角色以及角色之间的语义关系形成的高级语义块,视频事件语义的生成与用户对视频内容的关注角度直接相关。使用元组将事件表示为 Event = {eid, name, focus, ET, Attrs},其中:

① eid 是事件的唯一标识符。

② name 是事件名称。

③ focus 是生成该事件语义时所基于的关注角度。

④ ET = {t_i, t_j}是事件在视频中时间区间的起止时间记录。非复合事件的时间区间由事件角色的时间区间取并集得到,复合时间的时间区间由其子事件时间的区间取并集得到。

⑤ Attrs = {k_1: v_1, …, k_n: v_n}是可扩展的属性键值对。可以包含事件的语义时间信息、语义位置信息等。可根据需要进行扩展或补全。

以上定义的三类主体之间互相关联,共同构成完整的语义表示框架,从而表示丰富的视频语义。主体之间的关联方式的具体定义如下:

(1)语义对象间关联(Object-Object Relation):本文的方法着重刻画语义对象之间的事件无关的时空关联(Spatio-Temporal Relation)。时间关联主要是指两个语义对象在时间区间内的相对位置,如在相同的时间区间出现(equal)、在之前的时间区间出现(before)、在之后的时间区间出现(after)等。空间关联包括方向关联和拓扑关联,方向关联包括东(east)、南(south)、上(above)、下(below)等,拓扑关联包括覆盖(cover)、接触(touch)等。关于时空关联的详细定义见[21-23]。语义对象间的时空关联具有有向性,在图形框架中表示为一系列的有向边。使用元组将语义对象间的时空关联表示为 STRel = {type, oid1, oid2},其中:

① type 是时空关联的类型,比如"覆盖(cover)"。

② oid1 是指有向关系的起点语义对象的标识符。

③ oid2 是指有向关系的终点语义对象的标识符。

(2)语义对象 - 角色关联(Object-Role Relation):事件无关的底层语义对象在具体事件中实例化为事件中的角色,语义对象和角色间的关联称为实例化关联(Instantiation Relation)。在这种关联中,允许语义对象和语义角色之间存在一对一、一对多和多对一的数量关系。使用元组将实例化关联表示为 InsRel = {Os, Rs, eid},其中:

① Os = {oid1, …, oidn}是参与实例化的语义对象的标识符的集合。

② Rs = {rid1, …, ridn}是参与实例化的角色的

标识符的集合。

③ eid 是指实例化过程所面向的事件的标识符。

(3) 角色间关联(Role-Role Relation):角色的语义信息是与具体事件相关的,而角色的底层基础是语义对象,所以角色之间的关联是基于对象间的时空关联并结合具体事件的语义而形成的语义关联(Semantic Relation),角色间的语义关联也具有有向性。使用元组将角色间的语义关联表示为 SemRel = {type, rid1, rid2},其中:

① type 是语义关联的类型。该类型与具体的事件语义相关。

② rid1 是指有向关系的起点角色的标识符。

③ rid2 是指有向关系的终点角色的标识符。

(4)事件间关联(Event-Event Relation):本文的方法着重关注事件间的组合关联(Composition Relation),使用元组将事件间的组合关联表示为 ComRel = {Eid, Subs},其中:

① Eid 是复合事件的标识符。

② Subs = {eid1, …, eidn}是子事件的标识符的集合。

(5)事件 - 角色关联(Event-Role Relation):事件拥有角色,使用元组将事件和角色之间的关联(Owing Relation)表示为 OwReal = {eid, Roles},其中:

① eid 是事件的标识符。

② Roles = {rid1, …, ridn}是事件拥有的角色标识符的集合。

以上是本文提出的视频语义表示方法的框架定义。本文的方法关注对事件语义的多角度表示及相应的多种粒度划分方式的支持,在以上的定义中,事件的关注角度属性为事件语义的多角度拓展提供了基础,多种事件粒度的划分方式是通过每个不同角度下的事件的时间区间划分和事件间的组合关联实现的。为阐明上述实现过程,定义可扩展的对象 T = {T1, T2, …, Tn}。其中 T1 = {focus1, ETs}表示在关注角度为 focus1 时的事件粒度划分方式,ETs = {ET1, ET2, …, ETn}代表当前划分方式下的复合事件和非复合事件的时间区间集合。现有方法只支持单一角度下的单一划分方式,而在本文方法中,对象 T 的可扩展性即代表事件语义角度和事件粒度划分方式的多样性。

从图 3 可以看出,以某段视频为例,可将其多角度的事件语义表示和多种事件粒度划分过程形式化为对象 T = {T1,T2},其中 T1 = {focus1,{ET1,ET2}},T2 = {focus2,{ET1,ET2,ET3}},这代表两种角度下的两种

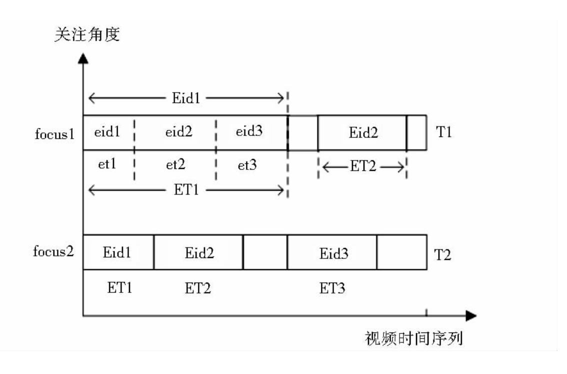


图 3　多角度下的多种事件粒度划分方式

粒度划分方式。图中角度为 focus1 时的 ET1 = {et1, et2, et3},表示事件 Eid1 与其子事件的组合关联,体现了事件的组合关联和时间区间划分的对应,说明了实现该过程的可行性。

在本文的视频语义表示方法中,视频语义的表示以底层的语义对象及其时空关联为基础,语义对象进一步实例化为具体事件中的角色,从而可复用地参与到多个具体事件的语义表示中。视频中的事件语义可以基于不同关注角度生成,在多个角度中形成多种粒度划分,事件的语义可以灵活地拓展和变化,事件涉及的角色及角色间的语义关联也随之变化,而框架底层的语义对象保持不变,只是基于不同的事件语义产生不同的实例化过程。实例化的角色是变化的高层事件语义和不变的底层语义对象之间相互联系的桥梁,这就是面向事件的视频语义表示方法的语义表示过程。下文将介绍本文方法的一些相关细节,并论述该方法是如何解决上一节所提到的视频语义表示中的关键问题的。

4.2　关键问题的解决

为支持自底向上的描述过程,本文定义的语义对象为使用当前的识别技术分析可得到的具有初级语义的语义对象,其属性都是与事件无关的。在对象间的关联方面则着重刻画与事件无关的时空关联。底层语义对象向上实例化为事件相关的角色,角色之间以底层语义对象的时空关联为基础,在事件中形成高层的语义关联,事件之间通过语义粒度的聚合再向上形成更高级的事件语义。这里举例说明语义对象的时空关联向上形成角色的语义关联的细节,如图 4 所示,两个底层的语义对象间的"时空位置接近(approach)"的时空关联,在具体的"抓捕"事件下,演变为"警察""罪犯"这两个角色间的"追逐(chase)"的语义关联。

在对视频事件的多角度解读的支持上,不同于过往研究中只支持先验的、单一角度的事件语义表示,本

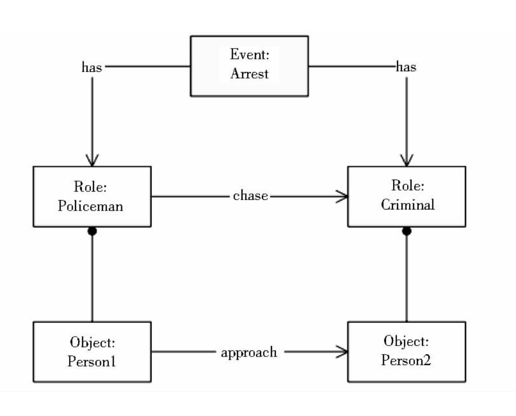


图 4 对象时空关联和角色语义关联

文的方法允许可扩展的、多个角度的事件语义表示。事件可通过上文中定义的“关注角度”(focus)属性进行角度区分,事件语义可从不同角度解读,每个角度下可产生不同的事件粒度划分方式。过往的研究方法一般只能刻画单一角度下的事件语义汇集的树状事件结构,而本文的方法使得最终的事件语义可以形成多角度汇集的网状事件结构。

在对事件粒度划分的支持上,本文设计了不同层次事件中的语义对象和角色的变化机制。语义对象和角色随事件变化的设计在于两个方面。一是在语义对象的数量上做切片,复合事件只关注其子事件中更能体现高层语义的语义对象,因此复合事件包含的对象集合是其所有子事件包含的对象集合的子集,比如在由子事件“Pass(传球)”、“Shoot(投篮)”组成的复合事件“Score(得分)”中,所有子事件涉及的语义对象共有 4 个,复合事件只涉及其中 2 个(见表 1);二是同一语义对象在不同层次的事件中使用不同的实例化角色,比如语义对象“Player1” 在子事件“传球”和复合事件“得分”中分别实例化为“传球者”角色和“助攻者”角色(见图 5)。

表 1 事件组合过程中的语义对象变化

语义对象信息	Pass, Shoot	Score
涉及的语义对象	Player1, Player2, Player3, Ball	Player1, Player2

在对自顶向下的语义补全的支持上,本文为语义对象、角色、事件定义了可扩展的属性键值对,可根据需要增加个性化的语义信息,能够满足自顶向下的语义补全的需求。并且由于存在角色作为语义对象和事件的中间层,在对具体事件中的角色进行事件相关的语义补全时,也不会改变对象本身的基础语义。

在对可扩展性和检索需求的支持上,首先,对可扩展性的支持体现在方法设计的各个方面,属性键值对的扩展、对象的实例化过程、事件的多角度生成方式分

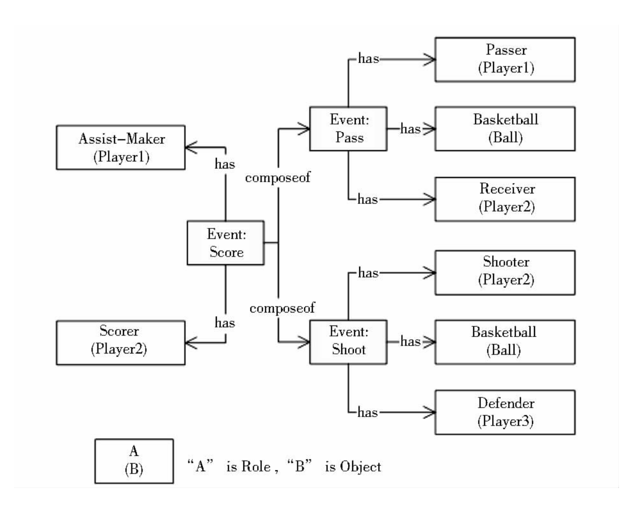


图 5 对象以不同角色参与多层次的事件

别支持了属性级、对象级、事件级的语义拓展。在检索能力方面,本文的表示方法除了可以支持语义丰富的图数据检索方式,还可以方便地基于语义对象或基于事件关注角度进行检索;可以通过实例化关系检索到对象实例化生成的所有角色。

4.3 方法应用实例

为体现面向事件的视频语义表示方法的应用效果,本文选取了一段篮球比赛视频的片段进行语义表示。

本文截取的视频片段来自于 CBA 的北京农商银行队(下称“A 队”)和四川品胜队(下称“B 队”)的一场比赛视频。本文截取的视频片段的连续关键帧如图 6、图 7 所示,其基本的场景信息为:在比赛即将结束时,A 队的球员“运动员 A3”传球给“运动员 A1”,“运动员 A1” 投篮命中,在比赛结束前将两队比分逆转。



图 6 视频关键帧一

李旭晖，吴青峰. 面向事件的视频语义表示方法[J]. 图书情报工作,2020,64(10):99－108.

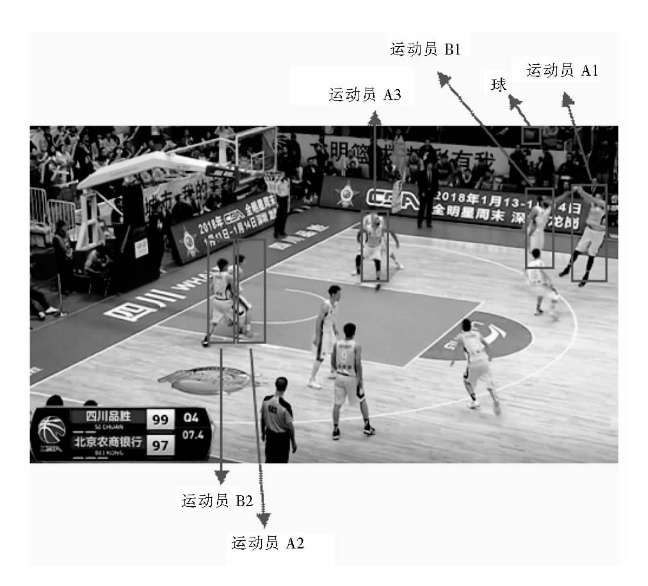


图 7　视频关键帧二

使用本文的方法对这段视频的语义进行表示，如图 8 所示，图中表示了这段视频的 3 个角度的事件语义，这 3 个角度分别以“持球人”“运动员 A2”“比分”为关注点。为方便展示，图中简化了主体属性的表示，将语义对象的名称、角色的语义标签、事件的名称和关注点属性直接展示在对应的矩形框中，其他属性未全部展示。图中未表示出所有的对象间、角色间关联，主要通过对象“运动员 A1”和“运动员 B1 ”及其对应的角色“投篮者”和“防守者”的关联体现了对象的时空关联到角色的语义关联的演化，图中两个语义对象间的时空关联“时空位置接近”在角色间演化为语义关联“拦截”。

在以“持球者”为关注点的语义表示中，对象“运动员 A1”“运动员 A3”和“球”分别实例化为“接球

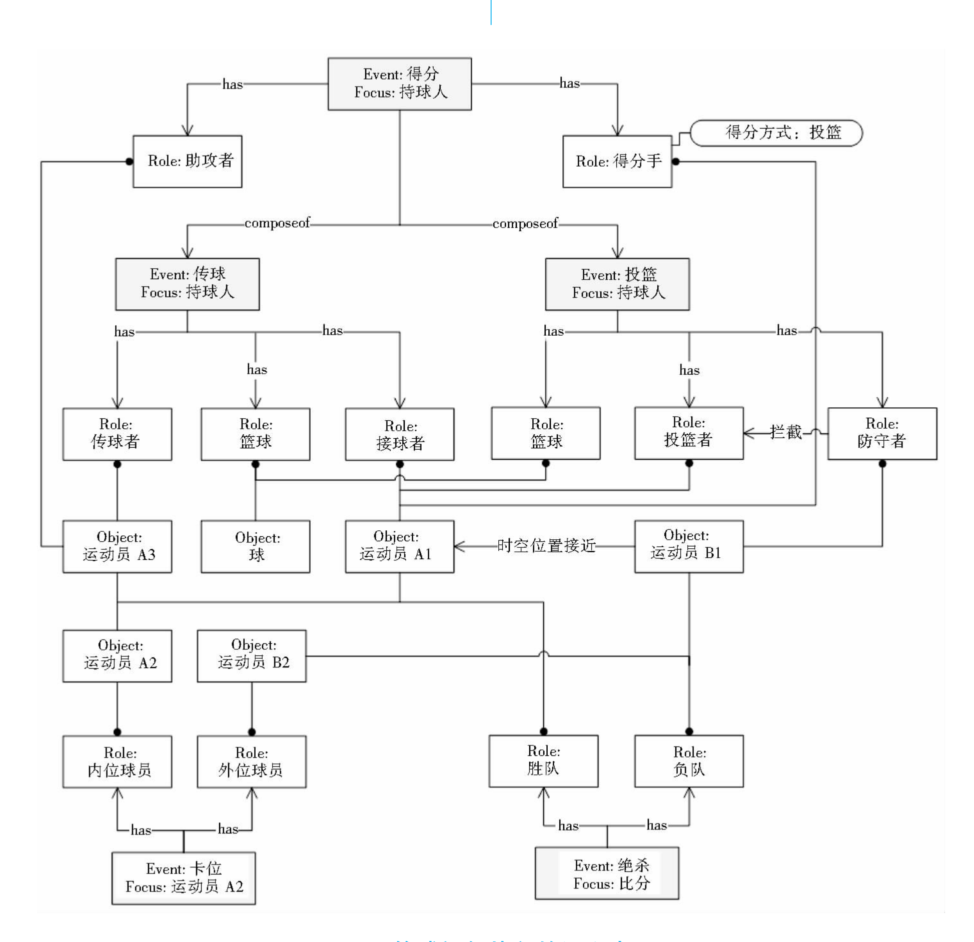


图 8　篮球视频片段的语义表示

者”“传球者”和“篮球”，成为参与“传球”事件的角色。对象“运动员 A1”“运动员 B1”和“球”分别实例化为“投篮者”“防守者”和“篮球”，成为参与“投篮”事件的角色。由于投篮命中，连续的“传球事件”和“投篮事件”组合成具有更大粒度的“得分”事件。在“得分”事件中，其子事件涉及的语义对象只有“运动员 A2”和“运动员 A1”与该层语义具有强相关性，所以“得分”件的语义表示只涉及了这两个对象，它们分别实例化为新的角色“助攻手”和“得分手”，后者在图中还展示了可扩展添加的属性键“得分方式”及其属性值“投

篮”。

以“运动员 A2”为关注点和以“比分”为关注点的事件语义的表示与上述过程类似,图中分别表示了以“运动员 A2”为关注点的“卡位”事件和以“比分”为关注点的“绝杀”事件。其中“绝杀”事件在篮球运动中是指在比赛将要结束前逆转比分并决定比赛胜负的比分事件,图中表示的参与绝杀事件的角色“胜队”和“负队”都是由多个语义对象以多对一的方式实例化得到的,这是本文的方法所支持的实例化机制。

以上 3 个角度的语义表示只用作本文的实例说明,在实际应用中,面向事件的视频语义表示方法支持扩展更多不同角度的事件语义表示。

4.4 相关方法比较与创新性说明

为了更好地理解各种视频语义表示方法的差异,并更直观地体现本文方法的优势,本节将本文方法与其他研究进行了比较,并对本文方法的创新性进行了说明。

本节选取了在相关工作中所提及的 6 种视频语义表示方法与本文方法进行比较,各类方法都具有基本的视频事件语义表示能力。本文方法主要针对现有方法中的事件语义表示角度和粒度划分方式单一、缺少灵活的对象语义变化机制的问题,因此在本节的比较中主要考察与之相关的以下 4 个需求:区分对象和角色、允许事件组合、支持多角度事件语义表示、具有对象语义变化机制。比较结果如表 2 所示,其中“√”代表方法能直接满足该需求或能够以类似的方式间接满足该需求,“×”代表方法不能满足该需求或没有定义相关的内容。

表 2 典型的视频语义表示方法比较

方法	区分对象和角色	允许事件组合	支持多角度事件语义表示	具有对象语义变化机制
AVIS[17]	√	×	×	×
VIDEX[18]	×	√	×	×
THVDM[20]	√	√	×	×
基于图模型的足球视频语义建模方法[13]	×	×	×	×
多层次地理视频语义模型[16]	√	√	×	×
视频分层语义联想模型[19]	×	√	×	×
面向事件的视频语义表示方法	√	√	√	√

在上述表格的几项指标中,本文提出面向事件的视频语义表示方法的效果最好。一方面是因为上述的一些方法在提出时还没有太多可以参考的相关工作,事件语义的复杂性还没有被关注,它们主要是在探索视频语义表示方法时厘清了事件相关的基本概念,为后来的研究提供了基础;另一方面,本文在前人的研究基础上聚焦于事件语义的多角度表示等面向事件复杂性的方面,对复杂事件语义表示中涉及的对象和角色的区分、事件组合和粒度划分、对象语义变化等方面进行了专门的考虑和设计。所以本文的方法在面向事件的语义表示时能够更契合需求。

具体而言,本文的方法具有以下创新:①具有完整的语义表示框架。本方法对视频语义的表示遵循自底向上的描述过程,在表示框架中涵盖了不同层次的语义信息,并将它们合理地关联了起来。②能够从多个角度表示事件语义。事件语义可以根据不同用户背景和需求从不同角度解读和生成,并产生多种事件粒度划分方式,可以形成多角度汇集的网状事件语义结构。③可以灵活地进行语义拓展。在本方法中,语义对象和事件具有低耦合关系,参与事件的语义对象的数量及其实例化角色的语义都有相应的变化机制。对象和角色的语义可随不同角度、粒度的事件语义的变化而灵活拓展。

5 结论

本文围绕视频语义表示的研究主题,提出了面向事件的视频语义表示方法,通过实例阐述了使用其进行语义表示的过程,并与现有相关研究进行了比较,说明了本文方法的创新之处。本文的方法解决了现有事件语义表示方法中事件语义表示角度和粒度划分方式单一、缺少灵活的对象语义变化机制等问题,并在视频语义补全和拓展方面提供了更好的支持。

无论是在公共领域的品牌传递、意识形态塑造[24]方面,还是在个人领域的知识学习[25]、消费娱乐[26]等方面,视频都正在成为日益重要的媒介。在实践中,面向事件的视频语义表示方法可以为视频数据资源的组织管理提供信息表示框架,能够支持用于满足用户精细化视频获取需求的系统的设计,可以为视频数据挖掘相关研究提供良好的中间数据结构。面向事件的视频语义表示方法可以具体应用于如下场景:①电子图书馆视频资源的组织管理。视频是当下读者获取信息的重要媒介,基于良好的语义表示方法重新组织视频资源,可以更好地发挥图书馆视频资源的价值和可用性[27]。②基于视频语义的检索或推荐系统的设计。

将视频语义表示方法应用于系统设计,可以为当前实际应用中的精细化的视频内容获取方案带来新的突破。③支持视频数据挖掘研究。结构化表示的视频语义信息能够支持涉及视频主题等高级语义相关的数据挖掘研究,并为其挖掘结果提供可解释性。

在后续的研究中,笔者将继续以下几个方面的工作:①语义表示工作的完善。本文强调在多角度事件语义表示下应当具有多种粒度划分方式,所以在事件间关联方面重点关注了与事件粒度划分最相关的组合关联,后续还可在多角度事件语义表示的基础上,在时序关联、因果关联等方面进一步完善。②语义数据模型的构建。笔者将为本文的视频语义表示方法建立通用的数据模型,并拟将其落实在基于图数据库的数据模式中。③系统设计和实现。使用纯人工标注的方法无法发挥模型的最大价值,笔者拟在通用数据模型的基础上,设计视频语义分析系统,实现自动或半自动的视频语义分析及语义信息的结构化表示和存储。

参考文献:

[1] CNNIC 互联网研究. 第 43 次 CNNIC 中国互联网报告发布[J]. 中国广播, 2019(4):48.
[2] 邓珞华, 邓东宁, 陈晟. 论视频图书馆的建设[J]. 大学图书馆学报, 2010, 28(2):70－73.
[3] 赵琨. 大数据环境下图书馆音视频资源发展及建设研究[J]. 图书馆建设, 2015, 248(2):64－68.
[4] 朱智贤. 现代认知心理学评述[J]. 北京师范大学学报,1985(1):1－6.
[5] 曹刘彬. 基于深度学习的图像及视频描述方法研究[D]. 太原:山西大学,2018.
[6] 周教生. 基于隐含语义分析的视频语义概念检测方法[J]. 信息通信,2018(2):141－143.
[7] 陈晨. 基于动作语义关联规则挖掘的视频分类研究[D]. 镇江:江苏大学,2018.
[8] VIJAYAKUMAR V, NEDUNCHEZHIAN R. Mining video association rules based on weighted temporal concepts[J]. ProQuest, 2012,9(4):297－303.
[9] LI G R, ZHANG W G, PANG J B, et al. Online web video topic detection and tracking with semi-supervised learning[J]. Multimedia systems,2016,22(1):115－125.
[10] 栾悉道,谢毓湘,韩智广,等. 新闻视频挖掘技术研究[J]. 计算机科学,2007,34(2):1－6.
[11] 王硕. 篮球视频精彩事件检测方法研究[D]. 西安:西安电子科技大学,2015.
[12] GUPTA A, WEYMOUTH T, JAIN R. Semantic queries with pictures: the VIMSYS model[C]//Proceedings of the seventeenth international conference on Very Large Data Bases. San Francisco: Morgan Kaufmann,1991:69－79.
[13] 王昊冉,白亮,老松杨. 基于图模型的足球视频语义建模方法[J]. 计算机科学, 2011,38(6):266－269,297.
[14] 张静,高伟,刘安安, 等. 基于运动轨迹的视频语义事件建模方法[J]. 电子测量技术,2013,36(9):31－36,40.
[15] 刘晓璐. 基于知识元的安防视频内容场景化表示及检索[D]. 大连:大连理工大学,2017.
[16] 谢潇,朱庆,张叶廷, 等. 多层次地理视频语义模型[J]. 测绘学报,2015(5):555－562.
[17] ADALI S, CANDAN K, CHEN S S, et al. The advanced video information system: data structures and query processing[J]. Multimedia systems, 1996, 4(4):172－186.
[18] TUSCH R, KOSCH H, BOSZORMENYI L. VIDEX: an integrated generic video indexing approach[C]//ACM international conference on multimedia. Los Angeles:ACM,2000:448－451.
[19] 刘宏哲,鲍泓,须德. 基于内容的视频分层语义联想模型[J]. 计算机应用, 2005,25(8):1797－1800.
[20] WANG Y, XING C X, ZHOU L Z. THVDM: a data model for video management in digital library[C]//Proceedings of the sixth international conference of Asian digital libraries. Berlin: Springer International Publishing, 2003:178－192.
[21] ALLEN J F. Maintaining knowledge about temporal intervals[J]. Readings in qualitative reasoning about physical systems, 1990, 26(11):361－372.
[22] LI J Z, OZSU M T, SZAFRON D. Modeling video temporal relationships in an object database management system[C]//Proceedings of the multimedia computing and networking. San Jose:SPIE, 1997:80－91.
[23] EGENHOFER, MAX J, FRANZOSA, ROBERT D. Point-set topological spatial relations[J]. International journal of geographical information science, 1991, 5(2):161－174.
[24] 朱旭. 挖掘短视频信息传播优势强化大学生意识形态教育[J]. 才智,2019(24):77.
[25] KILPATRICK C, STORR J, LIM K, et al. Exploring the use of entertainment-education YouTube videos focused on infection prevention and control[J]. American journal of infection control, 2018,46(11):1218－1223.
[26] 高士杰,吴丽丽,郭宸. 移动短视频广告创作与消费者心理研究[J]. 中国市场, 2019,(2):139－140.
[27] 陈春,李娜,马建霞. 国外图书馆非文本资源建设与服务现状分析及对我国的启示[J]. 图书情报工作,2015,59(10):53－59.

作者贡献说明:

李旭晖:提出选题,确定研究思路及论文框架,修改论文;
吴青峰:收集文献资料,撰写和修改论文。

Research on Video Semantic Representation for Events

Li Xuhui　Wu Qingfeng
School of Information Management, Wuhan University, Wuhan 430072

Abstract: [**Purpose/significance**] Video content is affecting the information life of a large number of people in China. The proper representation of video semantic is the key foundation for the current development of video content research and application. The existing methods of semantic representation of video only support the semantic representation of an event from one single perspective and lack the flexible change mechanism of relevant semantic objects, which results in insufficient semantic representation. So it is important to explore more effective video semantic representation methods. [**Method/process**] This paper proposed a video semantic representation method for events. This method considered the bidirectional nature of human cognitive processes and adopted a scalable way to support multi-perspective interpretation of event semantic. A change mechanism of number and semantic is designed to support relevant objects included in events. [**Result/conclusion**] This method has a complete semantic representation framework, which can effectively support multi-perspective interpretation of video events. It flexibly supports attribute-level, object-level, and event-level semantic extensions. Generally it can represent richer video semantics than existing methods.

Keywords: video semantic representation　multi-perspective　semantic extension